Geometric Approach To Segmentation And Protein Localization In Cell Cultured Assays

S. Raman, B. Parvin, C. Maxwell, and M.H. Barcellos-Hoff *

Lawrence Berkeley National Laboratory
Berkeley, Ca 94720

Abstract. Cell-based fluorescence imaging assays are heterogeneous requiring collection of a large number of images for detailed quantitative analysis. Complexities arise as a result of variation in spatial nonuniformity, shape, overlapping compartments, and scale. A new technique and methodology has been developed and tested for delineating subcellular morphology and partitioning overlapping compartments at multiple scales. This system is packaged as an integrated software platform for quantifying images that are obtained through fluorescence microscopy. Proposed methods are model-based, leveraging geometric shape properties of subcellular compartments and corresponding protein localization. From the morphological perspective, convexity constraint is imposed to delineate, partition, and group nuclear compartments. From the protein localization perspective, radial symmetry is imposed to localize punctate protein events at sub-micron resolution. The technique has been tested against 196 images that were generated to study centrosome abnormalities. Computed representations are evaluated against the ground truth annotation for comparative analysis.

1 Introduction

The response of tissues and biological material in general to exogenous stimuli is often heterogeneous and requires a large set of samples for each experimental variable, e.g., tissue type, type of stimuli, dosage, and concentration. These responses are often multidimensional and multispectral and can be imaged using different type of microscopy. Quantitative analysis of these responses is a necessary step toward visualization of large scale co-localization studies and construction of predictive models. Research in this area has spanned from learning techniques using texture-based features for characterizing patterns of protein expression [3] to geometric techniques using nonlinear diffusion [1, 12], curve evolution, and shape regularization for segmentation of subcellular compartments [4, 5, 12]. Often segmentation provides context for quantifying protein

* Research funded by the Low Dose Radiation Research Program, Biological and Environmental Research (BER), U.S. Department of Energy, under contract number DE-AC02-05CH11231 with the University of California. LBNL publication number is LBNL-58749. Points of contact: SRaman@lbl.gov and parvin@media.lbl.gov

expression. However when protein expression is not diffuse within a compartment, additional processing is needed within the specific context. This paper outlines a complete methodology and its evaluation for quantitative assessment of co-localization studies in cell culture assays. Although the technique has been tested against studying centrosomal abnormalities (CA), it is extensible to other phenotypic studies. As CA occur in less than 2% of normal tissue and in about 80% of breast cancers [9]. CA may serve as valuable prognostic and therapeutic targets. Various cellular stresses such as viral infection, exposure to ionizing radiation and altered microenvironmental stimuli, can augment the frequency and type of CA [8]. Within resting animal cells, the centrosome represents a major microtubule organizing center and is composed of a pair of centrioles and pericentriolar material. Prior to division, the centrosome will replicate during the DNA synthesis phase of the cell cycle. During mitosis replicated centrosomes will separate and nucleate a bipolar spindle that equally contacts and segregates the replicated genetic information into two daughter cells. One facet of CA refers to additional centrosomes (more than two), which leads to abnormal cell division. As CA are rare events in cell culture assays, large numbers of samples within and between treatment groups must be analyzed for objective results. Complexities arise as a result of nonuniform staining, overlapping nuclei, touching centrioles, and scales of these subcellular compartments. In the proposed system, these complexities are addressed through model-based techniques that are driven by the inherent geometries. These geometric constraints take advantage of the convexity features of the nuclear compartment and the radial symmetry of the centrosome. Nuclear extraction is initiated from differential spatial operators as opposed to intensity thresholding, which is a common practice in most ad-hoc solutions. These differential operators lead to edge fragments that are linked for high-level geometric analysis, partitioning, and grouping. Nuclear regions provide context for quantitative protein localization. When localization is not diffused, additional analysis is required to characterize punctate signals. These punctate signals may vary in shape, scale, and intensity. Furthermore, they often overlap and create additional complexity. These complexities are addressed through a special class of iterative voting, which is kernel-based, and its topography favors radial symmetries. It is robust with respect to variation in size and intensity, and delineates overlapped compartments.

Organization of this paper is as follows. Section 2 reviews previous research. Section 3 summarizes geometric segmentation of the nuclear regions which provide the context for protein localization. Section 4 outlines the spatial voting technique for protein localization. Section 5 provides (1)the experimental results for 196 images, and (2)the comparison of the system performance against manual analysis.

2 Previous work

The difficulties in localization of subcellular compartments are often due to variations in scale, noise, and topology. Other complexities originate from missing

data and perceptual boundaries that lead to diffusion and dispersion of the spatial grouping in the object space. Techniques for extraction of nuclear compartments are either through global thresholding or adaptive (localized) thresholding followed by watershed method for separating adjacent regions. Techniques in radial symmetries, as evident by centrosome configuration, can be classified into three different categories: (1) point operations leading to dense output, (2) clustering based on parameterized shape models or voting schemes, and (3) iterative techniques. Point operations are usually a series of cascade filters that are tuned for radial symmetries. These techniques use image gradient magnitudes and orientations to infer the center of mass for regions of interest [6, 7, 10]. Parametric techniques tend to be more robust as long as the geometric model captures pertinent shape features at a specific scale, e.g., Hough transform. Iterative methods, such as watershed [11], regularized centroid transform [12], and geometric voting Yang04, produce superior results because they compensate for larger variation of shape feaures.

The method implemented here falls into the category of iterative techniques which are adaptive to geometric perturbation and typically produce more stable results. This method shares several attributes with tensor-based voting [2], but it differs in that it is scalar and iterative.

3 Segmentation

In a typical 2D cell culture assay that is stained for nuclear compartment, some nuclei are isolated and others are clustered together to form clumps. Thus, the strategy is to detect isolated ones first, and then impose additional processing for the clumped regions. The image signature suggests that thresholding may be sufficient as an initial step; however, shading, nonuniform staining, and other artifacts demands a localized strategy. This localized strategy is an edge-based technique with a geometric convexity optimization approach for improved reliability. Edges are collected to form contours and then tested for convexity. If convexity fails then the clumped region is partitioned into multiple convex regions according to a geometric policy. Several intermediate steps are shown in Figure 1, and steps are as follows.

3.1 Boundary extraction and convexity

Let $I(x, y)$ be the original image with 2D image coordinates. An initial boundary is extracted by linking zero-crossing edges that are filtered by the gradient magnitude at the same scale. Zero-crossing (computed from Laplacian, $\nabla^2 I$) assures that boundaries are closed, and the gradient threshold assures that spurious contours are eliminated. Two gradient thresholds (low and high) are used to initiate linking from strong edges and fill the gaps with weak edge points. Next, each computed contour is approximated with a polygon and total angular change is computed to test for convexity. If the region is not convex then additional processing is initiated.

3.2 Grouping and Partitioning

Partitioning of clumped nuclei into distinct convex objects is through iterative decomposition and constraint satisfaction. Intuitively, these partitions should be terminated by folds in the boundary corresponding to positive curvature maxima. The main purpose of the constraint-based grouping is to limit the number of hypotheses and reduce computational cost. The net result of this process is a set of corresponding candidates for each positive curvature maxima point for potential decomposition. The following geometric constraints are enforced.

Positive curvature constraint The curvature at any point along the contour is given by $k = \frac{\delta^{'} x \delta^{''} y - \delta^{'} y \delta^{''} x}{(\delta^{'} x^2 + \delta^{'} y^2)^{3/2}}$. The contour derivatives are computed by convolving derivatives of a Gaussian with the contour information. The intent is to partition a clump of nuclei from the points of maximum curvature along the contour.

Antiparallel constraint The antiparallel constraint asserts that each pair of positive curvature maxima along the contour must be antiparallel, which is estimated by computing the tangent directions at each candidate point. This constraint reduces the number of hypotheses for a potential partition thus reducing the computational cost.

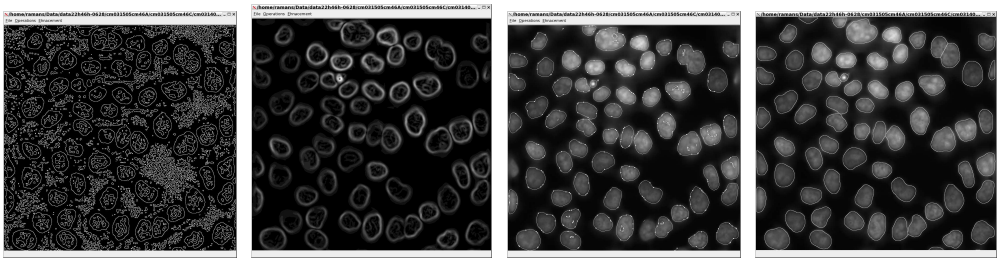
Non-intersecting constraint The Non-intersecting constraint asserts that a partition cannot intersect existing boundaries corresponding to the entire blob or other hypothesized partitions.

Convexity constraint The nuclear regions that occur in the cell culture are always convex, the convexity constraint enforces that the partition obtained has to be convex to avoid incorrect segmentation.

Grouping and partitioning Each clump is partitioned by linking pairs of positive curvature maxima that satisfy the above conditions. Each configuration has its own cost function, and the optimum configuration satisfies all the above mentioned constraints and will minimize $C = \Sigma_{i=1}^{n} \frac{\phi_i - \Pi}{\Pi}$, where n is the number of partition in a clump, determined by the system as follows. Essentially, the problem if reduced to grouping of curvature maximas in such a way that certain geometric constraints are satisfied.

Decomposition Algorithm

1. *Localize positive curvature maxima along the contour*
2. *Set initial number of compartments n:= 2*
3. *Construct a set of all valid configurations of n compartments by connecting valid pairs of positive curvature maxima satisfying the antiparallel, non-intersecting and convexity constraints*
4. *Evaluate cost of each configuration (per Equation 2)*
5. *Increment the compartment count n:=n+1 and repeat steps 3 and 4 until there is at least one configuration that has all convex compartments*
6. *Select the configuration with the least cost function*



(a) (b) (c) (d)

Fig. 1. Steps in segmentation: (a) Zero-crossing of Laplacian; (b) gradient image; (c) points of maximum curvature along contours; and (d) partitioning of clumped nuclei.

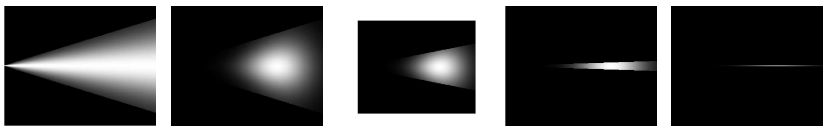
4 Protein localization

The problem of localizing punctate protein expression was first evaluated using Hough transform, cross correlation against training samples, and analysis of local intensity distribution. These Clustering based on Hough method proved to be scale sensitive, while correlation and intensity-based methods suffers from false positives and lack of geometric models. A geometric model is essential in the presence of scale varying and overlapping protein signals. A spatial class of spatial iterative voting is introduced to facilitate these requirements. Voting along gradient direction provides a hypothesis profile for saliency, e.g., punctate protein events. A specific kernel design (1) encodes the knowledge for saliency, (2) applied at each edge location along the gradient direction, and (3) refined and reoriented at each iteration step. The shape and evolution of these kernels, inferring center of mass, is shown in Figure 2. A brief review of the technique [13] is as follows: Let $I(x, y)$ be the original image, where the domain points (x, y) are 2D image coordinates. Let $\alpha(x, y)$ be the voting direction at each image point, where $\alpha(x, y) := (\cos(\theta(x, y)), \sin(\theta(x, y)))$ for some angle $\theta(x, y)$ that varies with the image location. Let $\{r_{\min}, r_{\max}\}$ be the radial range and Δ be the angular range. Let $V(x, y; r_{\min}, r_{\max}, \Delta)$ be the vote image, dependent on the radial and angular ranges and having the same dimensions as the original image. Let $A(x, y; r_{\min}, r_{\max}, \Delta)$ be the local voting area, defined at each image point (x, y) and dependent on the radial and angular ranges, defined by

$$A(x, y; r_{\min}, r_{\max}, \Delta) := \{(x \pm r\cos\phi, y \pm r\sin\phi) \mid r_{\min} \leq r \leq r_{\max} \text{ and } \theta(x, y) - \Delta \leq \phi \leq \theta(x, y) + \Delta\} \quad (1)$$

Finally, let $K(x, y; \sigma, \alpha, A)$ be a 2D Gaussian kernel with variance σ, masked by the local voting area $A(x, y; r_{\min}, r_{\max}, \Delta)$ and oriented in the voting direction

$\alpha(x, y)$. Figure 2 shows a subset of voting kernels that vary in topography, scale, and orientation.



(a) (b) (c) (d) (e)

Fig. 2. Kernel topography: (a-e)The Evolving kernel, used for the detection of radial symmetries (shown at a fixed orientation) has a trapezoidal active area with Gaussian distribution along both axes.

The iterative voting algorithm is outlined below for radial symmetry.

Iterative Voting

1. *Initialize the parameters:* Initialize $r_{\min}, r_{\max}, \Delta_{\max}$, and a sequence $\Delta_{\max} = \Delta_N < \Delta_{N-1} < \cdots < \Delta_0 = 0$. Set $n := N$, where N is the number of iterations, and let $\Delta_n = \Delta_{\max}$. Also fix a low gradient threshold, Γ_g and a kernel variance, σ, depending on the expected scale of salient features.
2. *Initialize the saliency feature image:* Define the feature image $F(x, y)$ to be the local external force at each pixel of the original image. The external force is often set to the gradient magnitude or maximum curvature depending upon the the type of saliency grouping and the presence of local feature boundaries.
3. *Initialize the voting direction and magnitude:* Compute the image gradient, $\nabla I(x, y)$, and its magnitude, $||\nabla I(x, y)||$. Define a pixel subset $S := \{(x, y)|\ ||\nabla I(x, y)|| > \Gamma_g\}$. For each grid point $(x, y) \in S$, define the voting direction to be

$$\alpha(x, y) := -\frac{\nabla I(x, y)}{||\nabla I(x, y)||}$$

4. *Compute the votes:* Reset the vote image $V(x, y; r_{\min}, r_{\max}, \Delta_n) = 0$ for all points (x, y). For each pixel $(x, y) \in S$, update the vote image as follows:

$$V(x, y; r_{\min}, r_{\max}, \Delta_n) := V(x, y; r_{\min}, r_{\max}, \Delta_n) + \sum_{(u,v)\in A(x,y;r_{\min},r_{\max},\Delta_n)} F(x - \tfrac{w}{2} + u, y - \tfrac{h}{2} + v)K(u, v; \sigma, \alpha, A),$$

where $w = \max(u)$ and $h = \max(v)$ are the maximum dimensions of the voting area.

5. *Update the voting direction:* For each grid point $(x, y) \in S$, revise the voting direction. Let

$$(u^*, v^*) = \arg \max_{(u,v) \in A(x,y;r_{\min},r_{\max},\Delta_n)} V(u, v; r_{\min}, r_{\max}, \Delta_n)$$

Let $d_x = u^* - x, d_y = v^* - y$, and

$$\alpha(x, y) = \frac{(d_x, d_y)}{\sqrt{d_x^2 + d_y^2}}$$

6. *Refine the angular range:* Let $n := n - 1$, and repeat steps 4-6 until $n = 0$.
7. *Determine the points of saliency:* Define the centers of mass or completed boundaries by thresholding the vote image:

$$C = \{(x, y) \mid V(x, y; r_{\min}, r_{\max}, \Delta_0) > \Gamma_v\}$$

(a) (b) (c) (d) (e)

Fig. 3. Detection of radial symmetries for a synthetic image simulating three overlapping centrosomes (a protein event): (a) original image; (b)-(e) voting landscape at each iteration.

An example of the application of radial kernels to overlapping objects is shown in Figure 3 together with the intermediate results. The voting landscape corresponds to the spatial clustering that is initially diffuse and subsequently refined and focused into distinct islands.

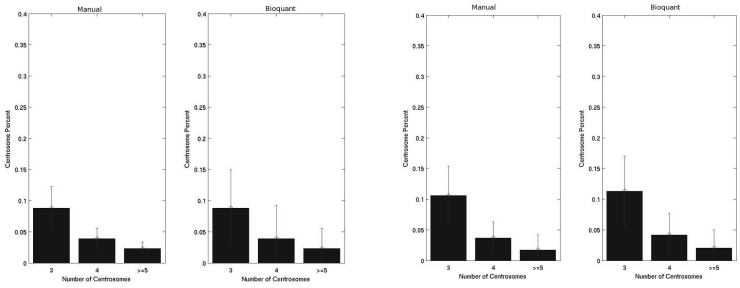
5 Experimental results and conclusion

A total of 196 images were processed to quantify number of abnormal centrosomes for each nucleus in the image. This result was then compared against manual count for validation, as shown in Figure 4. The system's error is at 1% and 10% for nuclear segmentation and quantitation of centrosome abnormality, respectively. Figure 5 shows the performance of the system on overlapping nuclear regions. It should be noted that in some cases there is no intensity decay

when adjacent nuclei overlap; watershed-based techniques can fail to produce proper decomposition of nuclear compartments under these conditions. In contrast, proposed geometric approach is invariant to intensity distribution as a basis for decomposition. An example of localization of centrosomes through voting is shown in Figure 6, where a rare event due to CA is captured in region 20 and region 45. Each punctate signal is assigned to the closest nuclear boundary.

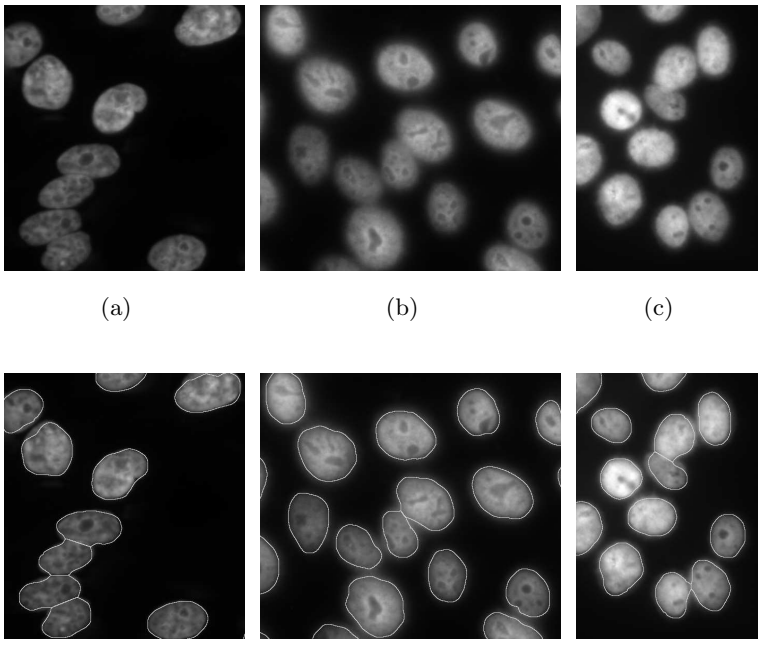
References

1. R. Malladi and J. Sethian. A unified approach to noise removal, image enhancement, and shape recovery. *IEEE Transactions on Image Processing*, 5(11):1554–1568, 1995.
2. G. Medioni, M.S. Lee, and C.K. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.
3. R. Murphy. Automated interpretation of subcellular locatoin patterns. In *IEEE Int. Symp. on Biomedical Imaging*, pages 53–56, April 2004.
4. C. et. al. Ortiz De Solorzano. Segmentation of nuclei and cells using membrane protein. *Journal of Microscopy*, 201:404–415, March 2001.
5. B. Parvin, Q. Yang, G. Fontenay, and M. Barcellos-Hoff. Biosig: An imaging bioinformatics system for phenotypic analysis. *IEEE Transactions on Systems, Man and Cybernetics*, 33(B5):814–824, October 2003.
6. D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *IJCV*, 14(2):119–130, March 1995.
7. D. Reisfeld and Y. Yeshurun. Preprocessing of face images: Detection of features and pose normalization. *CVIU*, 71(3):413–430, September 1998.
8. J. L. Salisbury. The contribution of epigenetic changes to abnormal centrosomes and genomic instability in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 6(2):203–12, April 2001.
9. J. L. Salisbury, A. B. D'Assoro, and W. L. Lingle. Centrosome amplification and the origin of chromosomal instability in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 9(3):275–83, July 2004.
10. G. Sela and M.D. Levine. Real-time attention for robotic vision. *Real-Time Imaging*, 3(3):173–194, June 1997.
11. L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *PAMI*, 13(6):583–598, June 1991.
12. Q. Yang and B. Parvin. Harmonic cut and regularized centroid transform for localization of subceullar structures. *IEEE Transactions on Biomedical Engineeing*, 50(4):469–475, April 2003.
13. Q. Yang and B. Parvin. Perceptual organization of radial symmetries. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 320–325, 2004.



(a) (b)

Fig. 4. Comparative results of abnormal centrosomes between manual and automated counting for two separate treatments. Each chart shows manual (on the left) and automated quantitation (on the right).



(a) (b) (c)

(d) (e) (f)

Fig. 5. Decomposition of overlapping nuclei: (a)-(c) original images; (d)-(f) decomposition results.

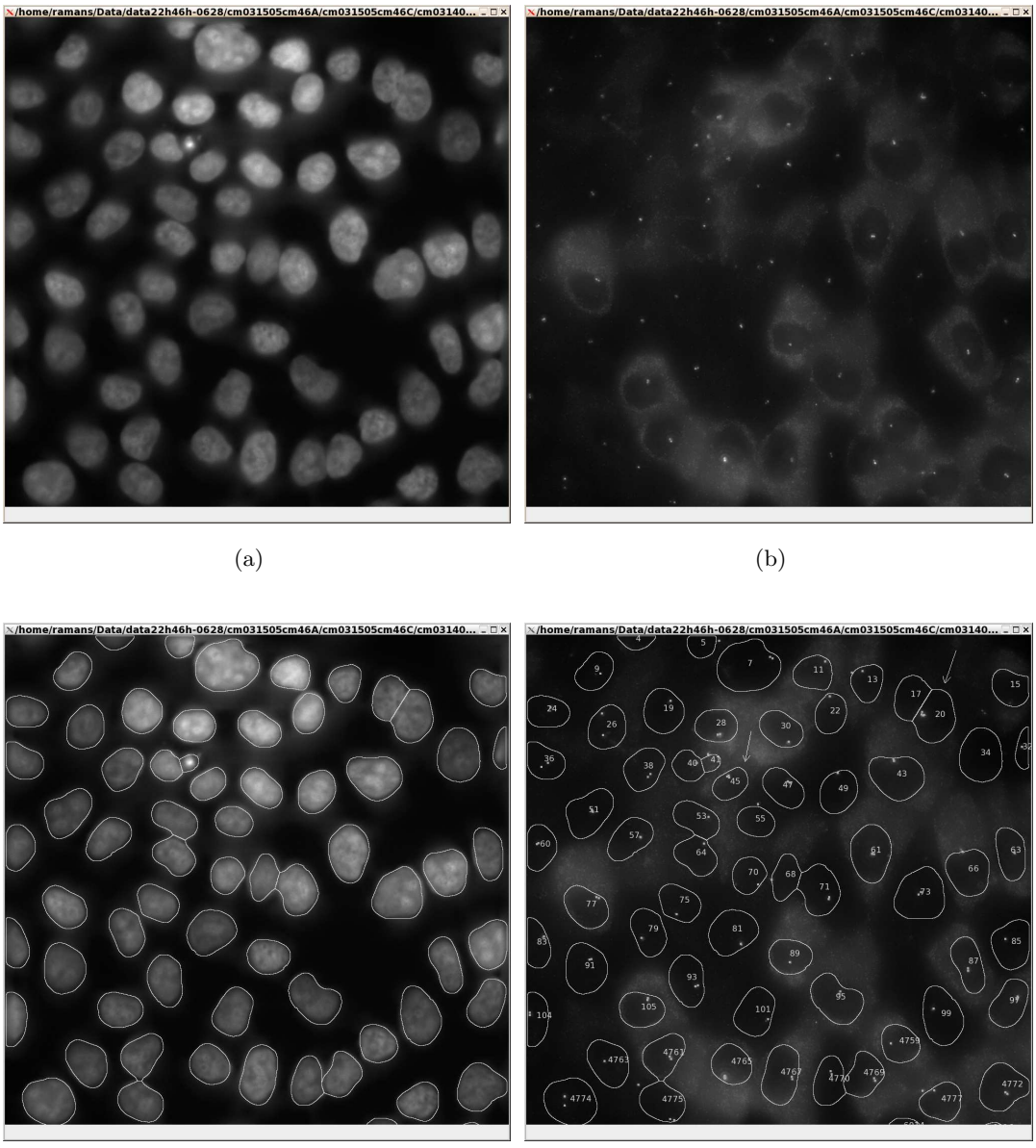


Fig. 6. Nuclear segmentation and centrosome localization indicates decomposition of overlapping nuclear compartments and detection of nearby punctate events corresponding to centrosome organelle: (a) original nuclear image; (b) corresponding centrosomes image; (c) segmented nuclear compartments; and (d) localized centrosomes. A rare event in nuclei 20 and 45 indicates four and three centrosomes, respectively. Nuclear and centrosome regions are represented by cyan contours and cyan dots, respectively. Ambiguities due to adjacent and overlapping regions (both nuclear and centrosomes) are resolved. Furthermore, pertinent events are measured in context. For example, centrosome abnormality of region 20 is referenced against correct nuclear size of morphology.